Chapter 2— Principles

Chapter 1 reviewed Verification, Validation, and Accreditation (VV&A) terminology and introduced some special topics that are of management-level concern with respect to VV&A efforts. This chapter delves a little deeper into the arcana of VV&A and discusses some principles that should be of concern to those tasked to develop and implement a VV&A program. The following principles are by no means the last word in VV&A orthodoxy, but they do distill the corporate memory of VV&A experts from government, industry, and academia into a convenient synopsis of hints that should help both the VV&A novice and the veteran avoid common pitfalls.

In the following paragraphs, 12 guiding principles of VV&A are proposed for your consideration and use. Some of these principles are no more than applied common sense; others require the nuance gained by experience to be appreciated. These principles form the basis for the whole concept of VV&A and for 76 recommended V&V techniques (described in Chapter 4) that can be used throughout the modeling and simulation (M&S) life cycle. Understanding and applying these principles is essential not only to the success of an M&S development effort, but also to the efficient application of VV&A resources and the credible integration of M&S into your specific application.

2.1 Principle 1

There is no such thing as an absolutely valid model.

By definition, a model is an *abstraction* or *approximate* representation of something. No model is ever totally representative (Banks *et al*., 1996, p. 407) and no model ever can be absolutely correct (Shannon, 1975).

As depicted in Figure 2-1 (Shannon, 1975; Sargent, 1992), an increase in model credibility infers a similar increase in model development costs. At the same time, the model utility also will increase but most likely at a decreasing rate. The point of intersection of these two curves is different for each M&S application.

Even at the extremes, however, the word *absolutely* must be interpreted in light of the application in which the model or simulation is to be used. A model can be *absolutely* invalid for one application but reasonable (within limits) for another. Thus, the same set of V&V data can be used to accredit a model for one application and yet point to the need for more development or additional V&V for another application.

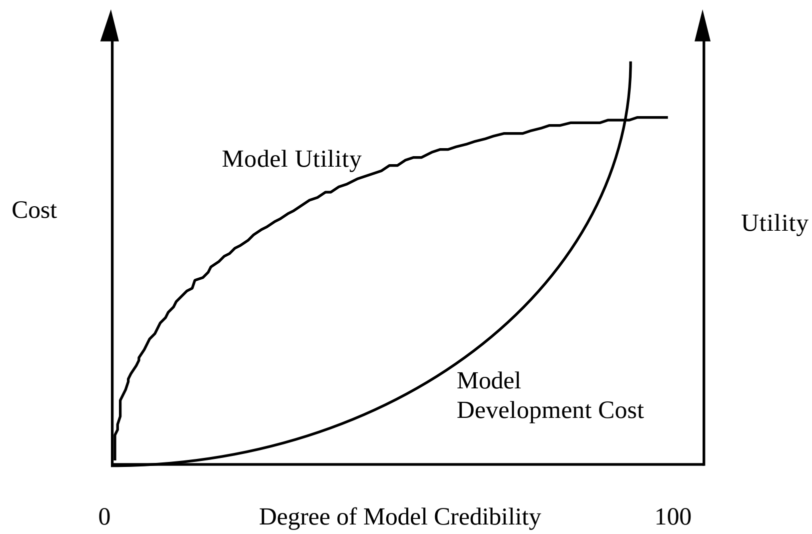

Figure 2-1. Model Credibility versus Cost and Utility

The goal of V&V is to exercise the model using a range of inputs, to identify and diagnose anomalous results, and to fix any problems encountered. Complete V&V would require the model or simulation to be tested and examined under all possible combinations of input conditions. Such combinations of input conditions easily can generate millions of logical execution paths that need to be checked for validity. Although some automated tools are available to assist in this task, time and budget constraints usually preclude exhaustive testing. As a result, the scope of V&V is usually tailored to the credibility requirements of the application. How much V&V is required depends on the intended use of the model or simulation. The greater the extent of the intended use, the more V&V will be required. It is impossible, however, to test all facets of a complex simulation.

2.2 Principle 2

VV&A should be an integral part of the entire M&S life cycle.

Despite the increased awareness of and attention to VV&A in recent years, many in the M&S community still refer to VV&A as if it were a single phase or step in the M&S development life cycle. The verb *to VV&A* is common among M&S practitioners, as are phrases such as "the model was VV&A'ed." Although M&S practitioners may know what they mean by these terms, their usage devoid of context has given many the incorrect impression that VV&A is a single task, rather than a process.

In reality, VV&A is a continuous activity performed throughout the entire life cycle of an M&S application, as shown in Chapter 3 (Figure 3-7). The sequential representation of the arrows in this figure is intended to show the general direction of M&S development, but the life cycle should not be interpreted as being strictly sequential. It is iterative in nature, and reverse transitions between phases of the life cycle are typical. Every phase of the life cycle has an appropriate set of VV&A activities whose breadth and depth is, in part, determined by the specifics of the application for which the simulation is being accredited. Deficiencies identified by VV&A may necessitate a return to an earlier development stage for revision and improvement.

For new M&S developments, conducting V&V for the first time in the life cycle only after the simulation has been developed completely is analogous to the teacher who gives only a final examination (Hetzel, 1984). No opportunity is provided throughout the semester to notify the student of serious deficiencies. Severe problems may go undetected until it is too late to do anything but fail the student. Frequent tests and homework throughout the semester are intended to inform students about their deficiencies so that they can take corrective action to improve their knowledge and performance as the course progresses.

The situation with VV&A is analogous. VV&A activities conducted throughout the M&S life cycle are intended to reveal deficiencies that might be present as development progresses from problem definition to the analysis of results. Tying VV&A to the M&S life cycle offers the opportunity to detect errors as early as possible and correct them at less cost and risk to the overall program. Too often there is a rush to implement a simulation, or "write the code." Sometimes simulations are built without formal specification of the requirements, resulting in a simulation that may not be relevant to the original problem to be answered by the use of this simulation. As a result, VV&A of the simulation becomes the only credibility assessment tool. And detecting and correcting the major modeling errors that VV&A will uncover at this stage is very time-consuming, complex, and expensive (Nance, 1994). *Correction of errors early in development always costs less than correction of errors later. If you are worried about the cost of VV&A, it is* *better to spend a little up front than a lot later.*

For legacy models and simulations, the situation appears a bit different on the surface, but in reality the same principle applies. Recall that by definition, legacy models and simulations were built before the development and implementation of consistent standards for VV&A. In practice, this has meant that little or no formal VV&A has been conducted for legacy models and simulations at any stage in their development life cycle. For this (and other) reasons, it has become common practice to use the term *legacy* as a pejorative modifier to "models and simulations." Recent postdevelopment VV&A experience with several widely used legacy models and simulations indicates that this practice is unjustified, however. For example, in the Susceptibility Model Assessment with Range Test (SMART) Project, several legacy models and simulations used by all Services to support acquisition and testing decisions were put through a rigorous V&V program with very positive results. This favorable reassessment is the result of efforts (a) by expert users to work out many of the problems associated with good legacy models and simulations over many years of use, (b) by frequent user group meetings, and (c) by reasonably well-disciplined configuration management practices developed through years of experience. The fact that such models and simulations fared well when subjected to the detailed scrutiny of formal VV&A performed after development should come as no surprise. In fact, the long period of use, modification, and consensus-building among the user community constitutes a form of face validation. (See Chapter 4 for a discussion of face validation as a V&V technique and its limitations.) Users, developers, and sponsors of the most frequently used legacy models and simulations have come to the realization in recent years that VV&A needs to be integral with development, and many have begun to incorporate formal VV&A into their development plans and budgets. Consequently, even for legacy models and simulations, it is becoming *de rigeur* for VV&A to be an integral part of the development life cycle.

2.3 Principle 3

A well-formulated problem is essential to the acceptability and accreditation of M&S results.

It has been said that a problem correctly formulated is half-solved (Watson, 1976). Albert Einstein once indicated that the correct formulation of a problem was even more crucial than its solution. The accuracy of the formulated problem greatly affects the accreditation and acceptability of M&S results. Insufficient problem definition and inadequate sponsor involvement in defining the problem are two of the most significant problems in the development and use of M&S. It must be recognized that, if problem formulation is poorly conducted, resulting in an incorrect problem definition, no matter how fantastically the problem is solved, the M&S results will be irrelevant. Balci and Nance (1985) present an approach to problem formulation and propose 38 indicators to use in assessing the formulated problem.

2.4 Principle 4

Credibility can be claimed only for the intended use of the model or simulation and for the prescribed conditions under which it has been tested.

Many people think that once simulation results have been compared with real-world data or test data, the simulation is "validated" for all time for all applications. Nothing could be further from the truth. The definitions for both validation and accreditation clearly require assessment of a model or simulation against a specific intended use. Global VV&A simply does not exist.

As stated in Principle 3, it is vitally important to the VV&A effort to define the intended use of the model or simulation clearly at the beginning of its application. (See Chapter 3 for an additional discussion of problem definition.) A well-defined problem statement will provide the necessary framework from which credibility of the model or simulation can be established and assessed.

Application objectives and requirements dictate how faithful the representation of a process, phenomenon, or system must be when compared with the real world for simulation results to be considered useful. Sometimes, 60 percent representation accuracy may be sufficient; sometimes, 95 percent accuracy may be required, depending on the type or importance of decisions that will be made based on the simulation results. Because the requirement for accuracy varies with the intended application, it is clear that a model's or simulation's credibility must be judged with respect to application-specific requirements and objectives. The adjective should always be used in front of terms such as credibility, validity, or accuracy to indicate that the judgment of validity has been made with respect to application-specific requirements.

Many factors can influence the intended use of a model or simulation. For example, modifications to simulations over time can introduce changes that require the results to be compared again with test data. In some cases, the simulation may have changed so much that the original validation data set has been rendered obsolete or inapplicable. Even if the data set is still acceptable, however, the simulation is still only valid over the range of conditions under which it is compared with test data. If the range of conditions is small, so is the range of validity.

Another factor that can influence the intended use of a model or simulation is the fact that the relationship between simulation inputs and outputs can be affected by the characteristics of the input conditions. The relationship that applies to one set of input conditions may produce absurd results when another set of input conditions is used. An intuitive example may help to illustrate this point. A model of a traffic intersection can be developed assuming a constant arrival rate of vehicles. (This is a reasonably valid assumption during the evening rush hour, for example.) When compared with actual traffic data at rush hour, the credibility of the model may be judged to be adequate with respect to evening rush hour conditions. The simulation will show erratic or invalid behavior when run under non-rush hour conditions, however. During this period, the arrival rate of vehicles is not constant, and a different simulation approach is required. Thus, the simulation could not be used to predict the behavior of a traffic intersection under non-rush hour conditions, even though it had been "validated" for rush hour conditions.

2.5 Principle 5

M&S validation does not guarantee the credibility and acceptability of analytical results derived from the use of simulation.

Model or simulation validity is a necessary, but not a sufficient, condition for the credibility and acceptability of the analytical results derived from use. Validity is judged with respect to the M&S objectives and requirements as they are defined. If the M&S objectives and requirements are incorrectly identified or the problem is improperly formulated, analytical results derived from the use of a model or simulation can be invalid or irrelevant; however, the model or simulation can still be found to be sufficiently valid by comparing it with the improperly defined system and requirements and with respect to the incorrectly identified objectives.

2.6 Principle 6

V&V of each submodel or federate does not imply overall simulation or federation credibility and vice versa.

Because a model is an abstraction of a system, with inherent assumptions, limitations, and approximations, it is unreasonable to expect perfect representation of all aspects of the modeled system. The credibility of each submodel or federate is judged to be sufficient with some error that is acceptable with respect to the M&S application objectives. Each submodel may be found to be sufficiently credible, but this does not imply that the whole M&S application is sufficiently credible. The allowable errors for the submodels may accumulate and become unacceptable for the whole model or simulation. Therefore, the whole model or simulation must be tested even if each submodel has been tested individually and found to be sufficiently credible. This same requirement applies to federations as it does to individual models and simulations.

Similarly, the determination that a federation has sufficient credibility cannot and does not imply that the individual federates are credible. Unfortunately, assumptions of federate credibility often are made to expedite the VV&A of a large federation. It must be remembered that submodel or federate V&V often rests on a previous or partial V&V effort that may not be relevant to the use of the model in the new application. The assumption that a model has undergone some level of scrutiny in the form of VV&A, when that VV&A actually has not been performed, leaves a significant gap in the understanding of the model and the resulting credibility assessment. When such assumptions are made on the VV&A of federate models, the investment of time and money on the VV&A of the federation results in a credibility assessment with little or no basis in reality and of limited use to the accrediting authority. Both the federation and its underlying federates and submodels must be verified and validated for the federation to be accredited for a specific use.

2.7 Principle 7

Accreditation is not a binary choice.

V&V results are not like the answers to Twenty Questions; very rarely are they simply "Yes, the model is good" or "No, the model is bad." Because a model is an abstraction of a system, with inherent assumptions, limitations, and approximations, it is unreasonable to expect perfect representation of all aspects of the modeled system when compared with test or other data. Consequently, it is more useful to consider the outcome of V&V activities in terms of a degree of confidence in M&S results that is expressed on a scale from 0 to 100, where 0 represents absolutely incorrect and 100 represents absolutely correct.

Figure 3-7 in Chapter 3 illustrates a range of responses available to the accrediting authority, including accredit the model, accredit the model with limitations to its use, modify the model, refer for additional V&V, or don't accredit the model. Accreditation agents, in particular, need to assist the accrediting authority in understanding his choices. Accreditation is not a fait accompli, as in the quick answer of “well, when it’s accredited,” which fails to acknowledge that the model may not be accredited. Neither is accreditation a yes or no decision. Chapter 5 provides additional amplification regarding the accreditation decision and is a must read for both accrediting authorities and program managers.

2.8 Principle 8

VV&A is both an art and a science, requiring creativity and insight.

Many people believe that because VV&A techniques are well-defined at the technical level, the integration of these techniques to establish model or simulation credibility for particular applications at least is similarly straightforward. This is not true. Cost-effective VV&A requires creativity and insight. It is not a checkmark in the box.

One must understand thoroughly the whole M&S application to design and implement effective tests and identify adequate test cases. Knowledge of the problem domain, expertise in the M&S methodology, and prior modeling and V&V experience are required for successful VV&A. It is not possible, however, for one person to understand fully all aspects of a large and complex model, especially if the model is stochastic and contains hundreds of concurrent activities.

The model or simulation developers are usually the most qualified to show the creativity and insight required for successful V&V because they are intimately knowledgeable about the model or simulation. They usually are biased, however, when it comes to testing their own models and simulations and, therefore, cannot be solely responsible for V&V. This limitation increases the difficulty inherent in V&V.

False beliefs exist, as indicated by Hetzel (1984) in referring to testing, beliefs that are often associated, however wrongly, with VV&A: testing is easy; anyone can do testing; no training or prior experience is required. As with testing, the difficulty of model or simulation V&V must not be underestimated. V&V must be well planned. It must be administered by the proponents and agents who are responsible for the model or simulation application. If V&V is delegated to contractors, oversight is required by the V&V agent. VV&A must also involve subject-matter experts to retain the focus on the needs of the warfighter. VV&A is a team effort.

2.9 Principle 9

The success of any VV&A effort is directly affected by the analyst.

The impact of the analysts who participate in the model or simulation application and, more particularly, in the VV&A of the model or simulation, is often overlooked. Analysts are key players in the use of M&S, from assisting in defining the problem to selecting the model or simulation to be used to running the simulation to interpreting the results. Military analysts frequently are called upon to provide the analytical perspective but also may serve as subject matter experts for applications involving warfare operations from their personal fields of expertise. Analysts can be found in virtually all roles within the M&S life cycle, including those of V&V agent, program manager, M&S proponent, and even accreditation authority.

Traditional training for analysts has not always included sufficient instruction in the use of M&S or the development of the technical skills required for VV&A. Part of the VV&A process itself, therefore, must consider the qualifications of the VV&A team, most notably, the analysts who are involved extensively in the M&S process. Emphasis must be placed on carefully achieving the right balance of education, experience, practical knowledge, and technical skills. The tendency to assume the qualifications of an analyst based on experience that is no longer timely or on an advanced degree should be avoided.

2.10 Principle 10

VV&A must be planned and documented.

Principle 2 stated that VV&A is not a phase or step in the M&S life cycle—it is a continuous activity. Ad hoc or haphazard V&V does not provide a reasonable measurement of model accuracy. Hetzel (1984) points out that "such testing may even be harmful in leading us to a false sense of security." Careful planning is required for successful VV&A efforts. Tests should be identified, test data or cases should be prepared, tests should be scheduled, and the whole VV&A process should be documented.

2.11 Principle 11

V&V requires some level of independence to minimize the effects of developer bias.

Many people have heard the term IV&V, where the "I" stands for independent. Some in the M&S community insist that V&V must be as independent as possible from the developer of M&S software to minimize the effects of bias: the fox must not be allowed to guard the chickens. It is certainly true that simulation testing and evaluation is more meaningful when conducted by someone who has no vested interest in the outcome. Model developers, who have the most knowledge of the model, may be the least unbiased in testing and evaluating their own products. This is understandable. Beyond the natural resistance to criticism, they may fear the repercussions that negative results may have on their performance appraisal, the credibility of their organization, and the prospect of future contracts. It is also true, however, that an insistence on excessive independence can lead to duplication of effort between the IV&V agent and the developer, to needlessly adversarial relationships, and to increased costs. It is possible, after all, to mistake ignorance for perspective. Neither extreme of independence is worth pursuing seriously, but no real hard and fast rules exist for how much "I" to put in IV&V between these extremes. The final decision must be derived ultimately from the trade-off between your budget and your level of confidence and trust in the model or simulation developer, as well as the requirements of your management chain to demonstrate independence.

2.12 Principle 12

Successful VV&A requires data that have been verified, validated, and certified.

The credibility of M&S results is related directly to the credibility of data used as input to or resulting from model use. Data need to be reviewed for accuracy and consistency, as described in Section 1.6.4. Guidelines are being developed to provide insight into the tools, techniques, processes, and procedures that assist the model user in determining data credibility.

References

Balci, O., & Nance, R.E. (1985). Formulated problem verification as an explicit requirement of model credibility. Simulation, 45(2), 76–86.

Banks, J., Carson, J.S., & Nelson, B.L. (1996). Discrete-event system simulation (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hetzel, W. (1984). The complete guide to software testing. Wellesley, MA: QED Information Sciences.

Nance, R.E. (1994). The conical methodology and the evolution of simulation model development. Annals of Operations Research, 53, 1–46.

Sargent, R.G. (1992). Validation and verification of simulation models. In J.J. Swain, D.

Goldsman, R.C. Crain, and J.R. Wilson (Eds.), Proceedings of the 1992 Winter Simulation Conference (pp. 104–114), Arlington, VA, 13–16 December 1992. Piscataway, NJ: IEEE.

Shannon, R.E. (1975). Systems simulation: The art and science. Englewood Cliffs, NJ: Prentice-Hall.

Watson, C.E. (1976). The problems of problem solving. Business Horizons, 19(4), 88–94.

Additional Reading

Balci, O., & Sargent, R.G. (1981). A methodology for cost-risk analysis in the statistical validation of simulation models. Communications of the ACM, 24(4), 190–197.

Balci, O., & Sargent, R.G. (1984). Validation of simulation models via simultaneous confidence intervals. American Journal of Mathematical and Management Sciences, 4(3&4), 375–406.

Johnson, M.E., & Mollaghasemi, M. (1994). Simulation input data modeling. Annals of Operations Research, 53, 47–75.

Law, A.M., & Kelton, W.D. (1991). Simulation modeling and analysis (2nd ed.). New York, NY: McGraw-Hill.

Schlesinger, S., et al. (1979). Terminology for model credibility. Simulation, 32(3), 103–104.